

Команди за филтриране и управление на данни

ас. Стоян Мечев

катедра „Информационни технологии“
ВВМУ „Н. Й. Вапцаров“

Четене и записване на файлове.

- Един от основните принципи на Unix и съответно на Linux е, да има много системни програми, всяка от които изпълнява опростена (концептуално) задача;
- Такива команди могат да се използват като строителни елементи в писането на собствени програми.
- Инструменти, които приемат входни данни от **stdin** и извеждат резултата на **stdout** се наричат филтърни команди или филтри.

• Командата **cat** (concatenate) –
предназначението ѝ е да обединява
(слепва) два или повече файла.

- ако не се укаже втори файл, резултатът се извежда на екран;
- на екран се извеждат всички файлове, независимо от типа им;
 - **reset** за възстановяване на терминала

Option	Result
-b	(engl. <i>number non-blank lines</i>) Numbers all non-blank lines in the output, starting at 1.
-E	(engl. <i>end-of-line</i>) Displays a \$ at the end of each line (useful to detect otherwise invisible space characters).
-n	(engl. <i>number</i>) Numbers all lines in the output, starting at 1.
-s	(engl. <i>squeeze</i>) Replaces sequences of empty lines by a single empty line.
-T	(engl. <i>tabs</i>) Displays tab characters as “^I”.
-v	(engl. <i>visible</i>) Makes control characters <i>c</i> visible as “^c”, characters <i>α</i> with character codes greater than 127 as “M-α”.
-A	(engl. <i>show all</i>) Same as -vET.

Разделяне на файлове на части - **split**

- `split [OPTION]... [FILE [PREFIX]]`
- Разделя FILE на части с имена PREFIXaa, PREFIXab, ...; размерът по подразбиране е 1000 реда, а PREFIX по подразбиране е 'x'. Ако FILE е пропуснат или е '-', прочита данните от стандартния вход
- Опции
 - `-b, --bytes=РАЗМЕР` – задава размер на изходните файлове. Приема различни мерни единици
 - `split -b 1K`
 - `-d` използва цифрови суфикси, започващи от 0, а не азбучни
 - `-- numeric-suffixes[=начална стойност]` същото като `-d`, но позволява задаване на началната стойност
 - `-l, --lines=N` разделяне N на брой реда(записа) за всеки на изходен файл
 - `-n, -- number=N` разделяне на N на брой изходни файлове
 - `ls /usr/bin|split -l 10`

Началото и края: head и tail

- **head** – показва началната част на файл (по подразбиране, първите 10 реда).
 - С опция `-n` се задава брой на редовете.
`ls -l /etc|head -n 15`
 - С опция `-c` се задава брой байтове от началото на файла. Може да се задават размерност – `b` – блокове, `k` – килобайтове, `m` – мегабайтове;
`head -c 1k some_file` ще изведе на екран първия килобайт на файла `some_file`.
 - Ако числото след `-c` е отрицателно, се показват всички, без последните байтове на файла. Например:
`head -c -20 some_file` ще покаже всички байтове, без последните 20 байта на файла `some_file`.
- **tail** – показва края на файл (по подразбиране, последните 10 реда).
 - С опция `-n` се задава брой на редовете.
 - Ако числото след `-n` започва с `+`, програмата извежда на екран всички редове от този ред нататък.
Например:
`tail -n +5 lines.txt` ще изведе съдържанието на файла `lines.txt` започвайки от ред номер 5.
 - С опция `-f` командата `tail` остава в режим на изчакване и актуализира информацията, ако се добавят нови редове към файла.
- тези команди могат да бъдат използвани като филтър с `|`
 - `dmesg|tail`

Сортиране на данни: **sort** и **uniq**

- **sort** – сортира данните, по предварително зададени критерии
 - използва последователността за сортиране за локализация на системата (LC_COLLATE). По подразбиране – от "a" до "Z".
 - ... **LC_COLLATE=de_DE.UTF-8** sort ще сортира данните според съответната кодировка.
 - Ползва полета "- k 2" – сортира по втора колона; "-k 2,4" сортира по втора и четвърта колона.
 - Ползва знакът за интервал като разделител.
 - Ако има няколко интервала – първият е разделител, останалите се приемат като стойност на полето.
 - sort -k2 text1
 - sort -h text2
- **uniq** – използва се за елиминиране на повтарящи се редове.
 - най-добре е, данните да са предварително сортирани;
 - приема само един файл за филтриране
 - uniq some_file **second_file**

Option		Result
-b	(blank)	Ignores leading blanks in field contents
-d	(dictionary)	Sorts in "dictionary order", i. e., only letters, digits and spaces are taken into account
-f	(fold)	Makes uppercase and lowercase letters equivalent
-i	(ignore)	Ignores non-printing characters
-k <field>[,<field'>]	(key)	Sort according to <field> (up to and including <field'>))
-n	(numeric)	Considers field value as a number and sorts according to its numeric value; leading blanks will be ignored
-o datei	(output)	Writes results to a file, whose name may match the original input file
-r	(reverse)	Sorts in descending order, i. e., Z to A
-t<char>	(terminate)	The <char> character is used as the field separator
-u	(unique)	Writes only the first of a sequence of equal output lines

Колони и полета **cut** и **paste**

- Тези команди са подходящи за текстови данни в табличен формат.
- **cut** – изрязва част от текста по колони съгласно зададените опции.
 - `cut -d: --output-delimiter=' ' -f 1,4` – относителни колони;
 - `cut -c3-7 text2` – абсолютни колони;
- **paste** – извежда на екран един или няколко файла в колони по различни начини.
 - Не е филтрираща команда;
 - Използва "ТАВ" като разделител
 - `paste text1 text2`

Благодаря за вниманието!

Въпроси?